Bayesi-Chain

Bayesi-Chain: Intelligent Identity Authentication

Juanita Blue, Joan Condell, Tom Lunney
Intelligent Systems Research Centre, University of Ulster, Derry, Northern Ireland, UK

Abstract— **In a bid to stamp out fraudulent crime, there is increased pressure on individuals to provide evidence that they possess a 'real' identity. Counterfeiting and fake identities have reduced confidence in traditional paper documentation as proof of identity, this has created a demand for an intelligent digital alternative. Recent government implementations and identity trends have also improved the popularity of digital forms of identification.**

Authentication of identity is a salient issue in the current climate where identity theft through record duplication is on the rise. Identity resolution techniques have proven effective in filtering duplicated and fake records in identity management systems. These techniques have been further improved by the implementation of machine learning techniques which are capable of revealing patterns and links that have formerly gone undetected. Research has also suggested that incorporating non-standard attributes in the form of social contextual data can increase the efficiency and success-rate of these fraud detection methods.

In the digital age where individuals are creating large digital footprints, online accounts and activities can prove to be a valuable source of information that may contribute to 'proof' that an asserted identity is genuine. Online social contextual data – or 'Digital identities' -- pertaining to real people are built over time and bolstered by associated accounts, relationships and attributes. This data is difficult to fake and therefore may have the capacity to provide proof of a 'real' identity.

This paper outlines the design and initial development of a solution that utilizes data sourced from an individual's digital footprint to assess the likelihood that it pertains to a 'real' identity. This is achieved through application of machine learning and Bayesian probabilistic modelling techniques. Where identity sources are considered reliable, a secure and intelligent digital identification artefact will be created. This artefact will emulate a blockchain-inspired ledger and may subsequently be used to prove identity in place of traditional paper documentation.

Keywords— identity; authentication; digital footprint, privacy; security

I. INTRODUCTION

The digital age has cultivated an environment where technology is relied upon to house and present the data that can verify an individual's identity. As individuals spend increasing amounts of time online, technology facilitates and records the activities that substantiate their lives. The digital age has also witnessed a rise in identity theft, where perpetrators use the identities of others for nefarious purposes and to essentially violate the law [1]. Duplicate and fake identities have become an important issue, as they present terrorists and criminals with the opportunity to commit various types of crime, while concealing their true identities [2].

Technology has ameliorated how identities are recorded and proven. A variety of attributes including name, birthdate, address and education contribute to proof of identity. However, supportive evidence is also required to verify that the individual is a true person [3]. Due to an increase in counterfeit documentation, there is no longer a pervasive confidence in paper documents as evidence of identity. There has been a catalytic shift toward the utilization of digital identification as a method of authentication. Innovative identification technology invoked by Estonia [4] has encouraged a futuristic trend where paper documents are no longer relied upon to attest to the identity of an individual.

Current implementations utilizing electronic records provide for convenience. However, entry processes that lack precision, verification and validation have caused fake, duplicate and erroneous records to become commonplace [5]. Subsequently, many identity schema lack the integrity to properly authenticate identities. Identity resolution can be invoked to determine if a single identity has been duplicated when described by variant data in separate records [6].

An individual's true identity is comprised of basic components, a personal identity represented by standard identifiers and also a social identity. A social identity refers to a person's biographical history that gathers over their lifetime [7] and is concerned with an individual's existence in the social context through interactions and social behaviours [8]. Research conducted in the area by Wang et al. has indicated that the use of non-standard attributes that relate to an individual's social behavior may contribute to the authentication or refutation of identities within identity resolution techniques [9][10]. In a cyber context, this social identity is demonstrated via an individual's digital footprint, meaning the activities conducted online and the traceable metadata that is shed with every online session. Therefore, it may be surmised that the same behaviours possess the potential to aid authentication of true identities.

This paper outlines the intelligent identity solution's initial design and development stages, where data sourced from an individual's 'digital footprint' is analysed through machine learning and probabilistic modelling to ascertain the likelihood of a 'true' identity. Where identities are deemed 'real', this data is combined with additional metadata to construct a secure and intelligent digital identification document. This intelligent digital identification could potentially allow individuals to assert their identity without requirement for traditional paper documents or electronic records.

II. BACKGROUND

In seeking a solution to prove identity through a digital footprint, various associated areas were explored. These included identity resolution, digital footprints, digital

978-1-5386-6046-1/18/$31.00 ©2018 IEEE

identification, machine learning, Bayesian probability, and blockchain.

A. Identity Resolution

Identity resolution is a process of semantic reconciliation that determines whether a single identity is the same when being described differently [8]. Conventional records consist of multiple attributes [9][10], identity resolution identifies if two records relate to one individual by comparing the content of individual corresponding fields [6] as depicted in Figure I. However, the accuracy of these attributes cannot be relied upon [5] and thus, they do not present a reliable source of information against which identity authentication can be performed. Identity resolution is used largely to detect identity theft and fraud.

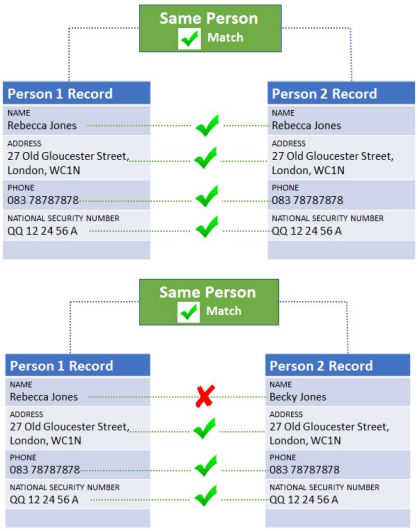

Figure I: Identity Resolution of Potential Duplicate Records

Existing resolution approaches are categorised as rule based or machine learned. Rule-based matching presents limitations in detecting true matches, especially in presence of missing values and erroneous data issues. Machine learned techniques are more efficient and automatically extract patterns and identify annotated matches. Distance/similarity measures between two records are defined for various attributes, resulting in the output of an over-all 'distance score'. Records are then matched if the score is above pre-determined threshold. Wang et al. developed an algorithm for detection of fake identities by comparison of several personal identifiers; combining them to produce a similarity score [5]. Fellegi et al. invoked probabilistic methods that predicted an estimated likelihood of record matches by estimating probabilistic parameters of the model via unsupervised techniques [11].

B. Social Contextual Information & Identity Resolution

Modern sociological literature indicates that two components form individual's identity: a personal identity and a social identity. An individual's personal identity is acquired from birth and includes identifiers such as name and date of birth; officially assigned identifiers such as a national security number (NSN); current physical descriptions such as height and weight and also biometric data such as fingerprints. A social identity is a person's biographical history, gathered over their lifetime [7], describing the social context of their life experience. Incorporating both these aspects allows for a more comprehensive understanding of identity.

In deviating from the utilization of traditional identifiers, an individual's social contextual information possesses attributes that authenticate their undeniable identity. Recent studies have recognized the value of social context data such as relationships and social behaviours in identity resolution. Identity matching through social behaviour and social relationship features was developed by Li et al. in 2010 [10]. Köpcke and Rahm also devised a categorical scheme that considered attribute-value-matchers that rely only on attributes that are descriptive and contextual matching to examine data gathered from social interaction links [12].

C. Digital Footprint

In the current world, individuals now possess two identities. A "real world identity", that is verified by official paper documentation, as well as a "digital identity", that is defined by an individual's use of the internet, including search history, online services, forums, blogs, and social media [13]. This use extends to create links to the real life identity of an individual. A digital footprint represents an individual's online presence and provides evidence of their digital and real world identities. It logs the trail and artifacts left behind by individuals interacting in a digital setting [14]. Digital footprints are persistent and link the past with the present, regardless of transitions and changes in an individual's life [15].

Online accounts provide many verified links to the attributes of real identities. Often these attributes are recounted across multiple accounts and sources. Almost every online account that is created requires an email address. Official online services require personal identifiers such as name, DOB, address and unique personal identifying numbers such as an NSN. Online shopping requires a postal address and payment details, searching the internet and the use of Google Maps often involves the use of an individual's current GPS location and potentially where they will be in the future and social media accounts represent confirmation of contacts, relationships [16] and professional and personal interests. Across several sources the same information relating to an individual is stored, reiterated and relied upon to conduct the simple tasks that form the operation of an individual's daily life.

A digital footprint is created unknowingly and with ease through automated logging such as the storage of cookies that has become an accepted aspect of being 'online'. However, it's direct descendent, a digital identity, in the social contextual form, is not so easily gained. The construction of a digital identity or reputation across multiple sources takes a significant amount of time to gather and its links to an individual's real world identity make it difficult to fake [9] as it is relied upon to conduct daily tasks. It requires multiple participants as it intertwines with external and official entities and it is bolstered by electronic records, email notifications, digital receipts, the lives of others and the metadata that forms components of the digital footprint and used to trace and record every online move.

As digital footprints map and record more and more aspects of an individual's real world life, they offer information and attributes that can be used to verify, validate and authenticate real identities, whilst also refuting those that are fake. These attributes include name, DOB, home address, phone number, email address, GPS locations, timestamps, financial information, professional affiliations, social relationships, personal health information, purchases, habits, interests and much, much more.

D. Digital Identification

Digital Identification refers to identity documents that are stored in electronic or digital format [17]. Well established as a technology trend, the demand for digital identification is increasing across institutions, corporations and even nations [18]. Organisations including the United Nations and World Bank ID4D initiatives have set a goal of providing everyone on the planet with legal digital identification documents by 2030 [19]. Estonia paved the way for the introduction of digital identification as part of its e-Estonia movement. The country pioneered the national use of online digital identification for all government transactions. This was achieved by issuing microchipped identity cards to all its citizens and residents [20].

E. Machne Learning

A branch of artificial intelligence, machine learning describes the type of techniques applied to detect meaningful patterns in data in an automated fashion. Machine leaning based technology exists in many facets of the modern world from email spam detection to market basket analysis. It is also widely applicable in various disciplines of science. These methods allow machines to emulate the ability of humans to learn from their observations and experience [21], creating models capable of predicting future outcomes, even where relevant data is absent [22].

F. Probability

"Probability is the branch of mathematics that studies the possible outcomes of given events together with the outcomes' relative likelihoods and distributions" [23]. Within this mathematical branch, Bayes Theorem/Rule provides a method of calculating the probability of an occurrence, when given the probability of another occurrence. Bayes Rule is used to relate conditionals of the format p(x|y) to the inverse, p(y|x) [24].

$$p(x \mid y) = \frac{p(y \mid x)p(x)}{p(y)} = \frac{p(y \mid x)p(x)}{\sum_{x'} p(y \mid x')p(x')}$$

In simple terms this is the basis for Bayesian Inference where Bayesian networks can calculate the chance of something occurring based other related information. Bayesian networks are commonly used in diagnostic systems where information is incomplete [25].

G. Blockchain

Blockchain initially made its technological debut as the backbone of digital currencies, Bitcoin in particular. The technology offers a secure, tamperproof ledger that records transactions while preserving the anonymity of the actors who conduct them. There are two types of records contained in a blockchain ledger: individual transactions and blocks. The initial block is comprised of a header and data that relates to transactions occurring within a set time period. A timestamp is added to the block and the combined data is hashed using one-way encryption [26].

Following creation of the first block, each subsequent block that is added to the ledger includes the previous block's hash and calculates its own hash. In order for any new block to be appended to the chain, its verification of its authenticity by a computational process called validation or consensus is required. When a block has been appended to the chain, it can be referenced in subsequent blocks, but cannot be altered. If there is an attempt to tamper with or remove a block, the hash values for previous and subsequent blocks will also change and disrupt the ledger's shared state. This prevents consensus from occurring, alerting the other nodes in the network of an error and preventing new blocks from being appended to the chain until the error is resolved [27].

III. Methodology

Bayesi-Chain aims to provide an intelligent method of secure and tamper-proof identity authentication. This is achieved by utilizing non-standard identity attributes sourced from an individual's digital footprint and encapsulating it in a blockchain inspired secure ledger. This solution is intended to be 'opt-in', meaning that individuals will submit their own data in order to potentially obtain a Bayesi-Chain Digital Identification document. Figure II depicts an overview of the design of the solution; a detailed description of each step follows.

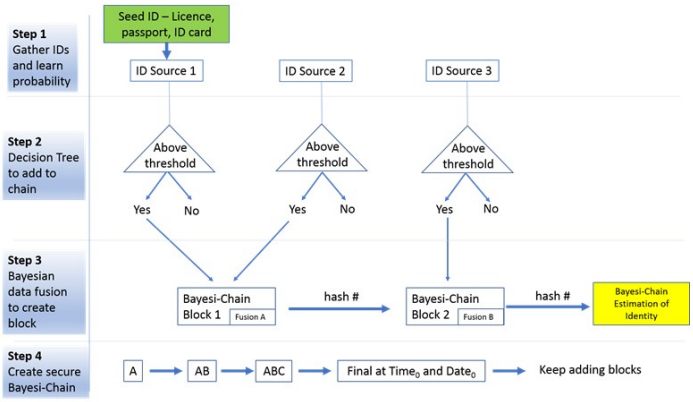

Figure II: Bayesi-Chain Design Overview

Combined methods of machine learning and probabilistic modelling will be utilized to estimate the reliability of identity sources pertaining to an individual's digital footprint; whereby weighted values associated with each of the attributes belonging to individual identity sources will produce an estimated score. Where the score is above an intelligently predetermined threshold, the combined standard and non-standard attributes will be fused into a 'block' with the current time and date stamp.

As additional identity sources are added to blocks, they will be combined with the previous 'block' by way of one-way hashing, producing a new hash value with each subsequent block. Further probabilistic modelling will be applied to estimate the overall 'Bayesi-Chain Estimation of Identity' score. An increased number of blocks in the chain results in a higher score. The higher the score, the higher the likelihood that the digital identification presented is 'real' and a valid authenticator of a true identity.

A. Step 1: Identity Sources & Identity Attributes

This section provides detail on Step 1 as depicted in Figure II. It categorises various identity sources associated with an individual's digital footprint and details the associated non-standard attributes. It describes how identity sources and attributes may be weighted and identifies how correlations and commonalities between attributes from multiple identity sources possess the capacity to verify personal details and authenticate identity. Furthermore there is an overview that provides explanation of extracting an estimated reliability score from identity sources and their attributes.

Identity Sources

An individual's digital footprint represents their online presence and is comprised of all their activities conducted on the internet. These activities span from cursory searches using engines such as Google or viewing a movie on Netflix, to more important tasks that facilitate the daily operation of life, such as communication by email or performing online banking tasks. It must be noted that any type of data associated with online accounts that is built up over time acts collectively as an efficient verifier of identity. The mere presence of data gathered over months and years is valuable, not necessarily the content of the data. For instance, an individual who wears a fitbit will gather data over extended periods including their heartrate, sleeping patterns, exercise regime and GPS locations. The existence of this extensive data that is also associated with other attributes such as their email address and mobile device including MAC address bolsters the probability that this data relates to a true identity.

Online activities provide varying degrees of valuable information. Based on the common requirement to create an online account or profile, certain sources may be considered more reliable than others. Online accounts such as online banking that previously required manual verification of paper identity documentation, or an online phone bill account that link to real-world information and payment card details may be considered more reliable than a social media account or subscription. Based on this premise, online accounts commonly utilized by individuals with an online presence have been categorised in Table I.

Table I: Identity Source Categories

Identity Source Categories	
Category 1: High Weight	
* Requires a paper document such as a passport that has been manually verified in order the create the online account	
• Online Banking	• Insurance Renewal
• Online Passport Application	• Online Voting
• Government Applications	• Motor Tax
Category 2: Medium Weight	
* Requires payment card information, provides verification for other accounts or links to standard attribute information such as phone number or residential address	
• Email Account	• Electricity Bill
• Online Shopping	• Online Prescriptions
• Phone Bill	• Mobile GPS tracking
Category 3: Low Weight	
* Requires little external verification and possesses weak, minimal or no links to real-world	
• Social Media	• Booking Engines
• Subscriptions	• Wearable Health Tracker

These online accounts (identity sources) form common elements of an individual's digital footprint and present evidence of real-world activities and links to information verifying the true identity of that individual. These links are through the unique subset of attributes that relate to each identity source.

Identity Attributes

Table II outlines the potential subsets of attributes that may be available from various, commonly held identity sources.

Table II: Attribute Subsets Available From Identity Sources

Identity Sources \ Identity Attributes	Name	Address	Phone Number	DOB	NSN	Email Address	Relationships	Habits	Financial
Facebook	Maybe	Maybe	Maybe	Maybe	No	Yes	Yes	Yes	No
Fitbit	Yes	Maybe -GPS	No	Yes	No	Yes	Maybe	Yes	No
Email	Yes		Maybe	No	No	Yes	Yes	Maybe	No
Phone Bill	Yes	Yes	Yes	No	No	Yes	Maybe	Maybe	Yes
Amazon - Shopping	Yes	Yes	Yes	No	No	Yes	No	Yes	Yes
Netflix	Yes	Maybe	Maybe	No	No	Yes	Maybe	Yes	Yes
Google	Yes	Maybe -GPS	Maybe	No	No	Yes	Yes	Yes	No
Online Banking	Yes	Yes	Yes	Yes	Maybe	Yes	Maybe	Yes	Yes

Attributes provided by various identity sources can provide valuable links that verify information pertaining to an individual's true identity, in addition they provide validation of standard attributes such as residential address and phone number, Table II provides examples of these links.

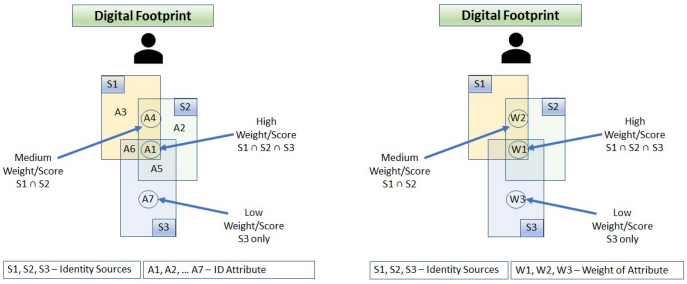

Figure III: Identity Sources with Common Attributes

The unique set of attributes possessed by each source can be weighted based on the categorised reliability of the source. This will determine the weight that may be applied to each attribute. Table II also demonstrates that Category 1 and Category 2 identity sources typically possess higher volumes of more salient attributes that link to an individual's real-world identity.

Many attributes correlate between identity sources as shown in Figure III, improving their reliability. Figure IV shows where attributes are found to be common across several identity sources, their score can subsequently be increased as their frequency increases the likelihood that they are associated with a true identity.

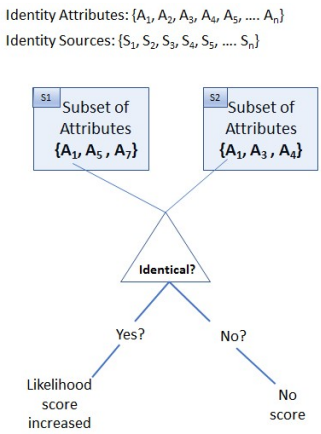

Figure IV: Common Identity Attributes

Identical attributes that are repeatedly associated with an individual's identity will be bolstered by their common use. These attributes will experience an increase in score, while

attributes not verified by alternate data sources will gain no additional score. An example of an attribute that is common across several identity sources is an email address linked to multiple online accounts.

Weighting of Identity Sources and Identity Attributes

Initially standard weights sourced from 'Police Vetting' forms may be applied to input the reliability of various identity sources, based on the subset of attribute types they possess. Weighting of identity attributes will be intelligently calculated using machine learning techniques. This calculation will be based on the weight of the identity source, the type of attribute and the value's frequency of presence across multiple identity sources.

Calculating Estimated Probability Values for Identity Sources

To calculate the estimated reliability score for each identity source, Bayesian probabilistic modelling will be applied to the associated subset of attributes and their learned weights. This will output an estimated value for the reliability of the identity source. As the reliability of further identity sources is estimated, machine learning will again be invoked to identify a suitable threshold to determine if an identity source is considered reliable.

B. Reliability Estimation Threshold & Decision Tree

This section provides detail on the process involved in Step 2, depicted in Figure II. Inclusion of identity sources within the final Bayesi-Chain digital identification document will be dependent on the intelligently determined threshold.

Identity sources that produce a reliability estimation score below the threshold will be discarded. Those that produce a reliability estimation score above the threshold will move to Step 3 for data fusion and hashing as depicted in Figure V.

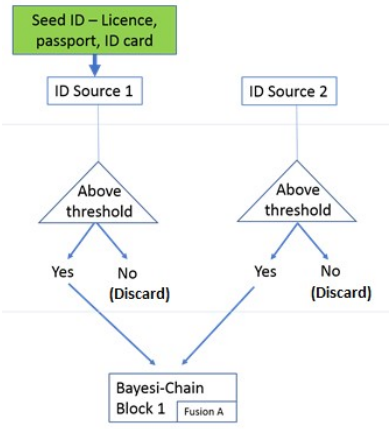

Figure V: Reliability Estimation Threshold & Decision Tree

C. Data Fusion, Block Creation & Hashing

This section details Step 3, depicted in Figure II. It provides an overview of the data fusion and block creation processes that are conducted where an identity source has scored above the threshold and has been intelligently determined as a reliable identity source.

Data Fusion

For this process to be initiated there must be a minimum of two identity sources that are determined to be reliable. One of these must be a 'seed' or Category 1 identity source, where manual verification of a paper identity document was required to create the online account.

Where two suitable identity sources have successfully undergone Steps 1 & 2, the information from both sources will be combined in a block using data fusion techniques. This block will also be time and date stamped. Bayesian probability calculations will be applied to the estimated reliability scores gained from the identity sources. Additionally, attributes which are common across both identity sources, will gain an increase in weight. This will produce a score that represents the 'Bayesi-Chain estimation of identity'.

As additional identity sources are added to subsequent blocks, they too will be fused with the information from previous blocks, again seeking common attributes across multiple identity sources and with consideration for the reliability scores of the identity sources.

Block Creation

The initial block that is created will store the identity sources and their associated attributes, the estimated reliability scores, a date/time stamp and the Bayesi-Chain estimation of identity. This data will be hashed using a one-way encryption mechanism to produce a fixed length hexadecimal string.

Each subsequent identity source that is added to the chain will be combined in a new block with the previous block's identity source and attribute information, a date/time stamp and the previous block's hash value. Thus, producing a blockchain inspired ledger of all the identity sources and their attributes.

Hashing

One-way encryption (hashing) is incorporated with the addition of each new block. This ensures that the information stored in each block is tamperproof, as a change to a single character in the block will result in an entirely different fixed length hexadecimal string. The inclusion of each block's hash within the next block facilitates the ability to pinpoint any block that has been tampered with, as the subsequent hash values will not match. Date and time stamp information also ensure that the block's point of creation is recorded and that each block, and therefore each hash, will be entirely unique. This hash/chain method is similar to those included in various blockchain technologies.

Create Secure Bayesi-Chain Digital Identification

Following successful data fusion, Bayesian probabilistic calculations and hashing of multiple blocks, Bayesi-Chain digital identification will begin to form. This identity will be represented by a unique identifying string that is linked to an individual's overall Bayesi-Chain estimation of identity score.

As subsequent blocks are added to the chain, the Bayesi-Chain estimation of identity score will increase. As the Bayesi-Chain estimation of identity score increases, so does the likelihood that the digital identification represents a true identity.

It is important to note that subsequent blocks may need to be added to Bayesi-Chain in a periodical manner to ensure the information contained therein remains recent and current. Bayesi-Chain digital identification scores should experience a decrease where blocks are not added within a set timeframe.

It is envisaged that the Bayesi-Chain digital identification document, verified by the estimation score and unique identifying string, could potentially be displayed via an interface on a mobile device. Possibly with the inclusion of a QR code, the interface could be used to authenticate identity in place of traditional paper documents. The digital verification

solution could also be invoked as an authenticator for official online application forms and communications.

IV. Development Progress & Results

This paper outlines the design and initial development stages of the aforementioned intelligent digital identification solution. Currently, an initial test environment has been created in the form of a relational database which includes both 'real' and 'fake' identities. This synthetic database is comprised of a core table that stores traditional personal identifiers including full name, date of birth, gender and home address. Each record's primary key links the records to a secondary table that documents relevant online accounts that have been nominated as potential identity sources. This table subsequently links to several other tables that store the attributes associated with each online source.

Preliminary execution of identity resolution techniques that incorporate the additional social contextual data have proven successful in discerning between identities that are likely 'fake' and likely 'real'. These tests were conducted using standard prepared statements.

The next phase of the development will require researchers to apply machine learning techniques and Bayesian probabilistic modelling to data sources and attributes in order to intelligently determine reliability scores for identity sources, weight attributes and determine appropriate thresholds.

V. Conclusion

This paper has identified the importance of effective identity verification and authentication in preventing criminal and terrorist activity facilitated by fraudulent identities. It highlights methods by which fraudulent identities are gained and purported for nefarious purposes by those who intend to commit further illegal acts. It also highlights the necessity for law abiding citizens to protect and prove their own real identities.

Identity resolution methods that seek to authenticate or refute similar or duplicate identities have been reviewed. This emphasizes that resolution based on standard personal identifiers has extreme limitations, providing substantial evidence that integration of an individual's social contextual data can greatly improve the success of identity resolution when paired with machine learning and Bayesian probabilistic modelling techniques. The inferred conclusion is that the same social contextual data could be used to authenticate a 'real' identity.

This paper outlines the design of an algorithm that aims to authenticate or refute identities based on information gained from digital footprints. It aims to demonstrate the feasibility of the concept, documenting the flow of each integrated step in the creation of the final digital identity document. Initial tests that have been executed have been successful in contributing to this end. The future work documented will continue to test the efficacy of the Bayesi-Chain Intelligent Identity Authentication Solution.

References

[1] Spalevic, Z. and Ilic, M., "The use of the dark web for the purpose of illegal activity spreading", ЕКОНОМИКА, Vol. 63, January-March 2017

[2] Kean, T.H., Kojm, C.A. Zelikow, P., Thompson, J.R., Gorton, S., Roemer, T.J., Gorelick, J.S. , Lehman, J.F., F.F. Fielding, F.F., Kerrey, B.,"The 9/11 Commission Report" (2004). URL: http://govinfo.library.unt.edu/911/report /index.htm

[3] Blue, J., Condell, J., "Identity Document Authentication using Steganographic Techniques: The Challenges of Noise", Ulster University, ISSC 2017, Kilarney, Ireland, (2017).

[4] Korjus, K., "Estonia is enhancing the security of its digital identities" (2017) URL: https://medium.com/e-residency-blog/estonia-is-enhancing-the-security-of-its-digital-identities-361b9a3c9c52

[5] Wang, G.A., H.C. Chen, H.C., Xu, J.J., Atabakhsh, H., "Automatically detecting deceptive criminal identities", Communication, ACM 47, page 70 –76 (2004).

[6] Christen, P. and Winkler, W., "Record Linkage", Encyclopedia of machine learning and data mining, Springer (2017).

[7] Vignoles, V.L., "Identity: Personal AND Social". University of Sussex, Oxford Handbook of Personality and Social Psychology, Second edition, Oxford University Press, London (2017).

[8] Hogg, M., Abrams, D., Brewer, M., "Social identity: The role of self in group processes and intergroup relations", Sage Journals, Vol 20, Issue 5 (2017).

[9] Wang, G.A., Chen, H.C., Xu, J.J., Atabakhsh, H., "Automatically detecting criminal identity deception: an adaptive detection algorithm". IEEE Transport Systems Management, Part A-Systems Humans 36, page 988–999 (2006).

[10] Li, J., Wang, G.A., Chen, H., "Identity matching using personal and social identity features", Information Systems Frontier 13, page 101–113 (2010).

[11] Fellegi, I.P., and Sunter, A.B., " A theory for record linkage", American Statistics Association, 64, 1183–1210 (1969).

[12] Köpcke, H., and Rahm, E., "Frameworks for entity matching: a comparison". Data Knowledge Eng. 69, page 197–210 (2010).

[13] Park, M., "AR is on the verge of transforming the human-computer relationship". VB, October (2017). URL: https://venturebeat.com/2017/10/30/ar-is-on-the-verge-of-transforming-the-human-computer-relationship/

[14] Fish, T., "My digital footprint". AMF Ventures Limited, Futuretext, London (2009).

[15] Haimson, O.L., Brubaker, J.R., Dombrowski, L., Hayes, G. "Digital footprints and changing networks during online identity transitions ", Dept of Infomatics, University of California, Irvine, CA, Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, Pages 2895-2907, San Jose, CA (2016).

[16] Xiang, R., Neville, J., Rogati, M.,"Model relationship strength in social networks", Purdue University, West Lafayette, IN, Proceedings of the 19th international conference on World wide web, Pages 981-990, Raleigh, NC (2010).

[17] Spinks, R., "Do we trust digital identification", The Guardian, 25 July (2016) URL: https://www.theguardian.com/media-network/2016/jul/25/do-we-trust-digital-identification

[18] Gemalto, "Digital Identity Trends: Five forces that are shaping 2017", (2017) URL: https://www.gemalto.com/govt/identity/digital-identity-trends

[19] World Bank, "Identification for Development (ID4D) Annual Report 2017", World Bank Group, Page 4, (2017).

[20] Steckman, L.M. and Andrews, M.J., "Online around the world", ABC-CLIO LLC, Santa Barbara, Page 85, (2017).

[21] Furey, E., "HABITS: A history aware based indoor tracking system", Faculty of Computing and Engineering, Ulster University, Londonderry, (2011).

[22] Bell, J., "Machine Learning: Hands on for developers and technical professionals". John Wiley and Sons Ltd, Indianapolis, pages 20-21(2015).

[23] Weisstein, E.W., "Probability". Mathworld, Wolfram Alpha (2017). URL: http://mathworld.wolfram.com/Probability.html

[24] Bernado, J.M., Smith, A.F.M., "Bayesian Theory", John Wiley & Sons, Chichester, (2000).

[25] Fox, D., Hightower, J., Kauz, H., Liao, L. & Patterson, D.J., "Bayesian techniques for location estimation", Proceedings of the 2003 Workshop on Location-Aware Computing, pp. 16. Seattle, Washington, USA, (2003).

[26] Bashir, I., "Mastering Blockchain", Packt Publishing, Birmingham, page 18 (2017).

[27] Mougaya, W., "The Business Blockchain: Promise, practice and the application of the next internet technology", John Wiley & Sons, Hoboken, NJ, page 7, (2016).